CORRÉLATION CANONIQUE (K6)
(05 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

Une **corrélation canonique** est une corrélation définie en **analyse canonique**.

(i) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé** et $\zeta = (\xi, \eta) : \Omega \mapsto \mathbf{R}^K \times \mathbf{R}^G$ un couple de **vecteurs aléatoires** de carré intégrable. On pose (**matrice des covariances** théoriques) :

$$(1) \qquad \Sigma = V\,\zeta = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi\eta} \\ \Sigma_{\eta\xi} & \Sigma_{\eta\eta} \end{pmatrix},$$

cette matrice étant partitionnée conformément à ζ.

Par définition, les (carrés des) **coefficients de corrélation canoniques** (théoriques) entre ξ et η sont les solutions $\rho_1^2, ..., \rho_K^2$ de l'équation en ρ^2 :

$$(2) \qquad \text{Dét}\,(\rho^2\,\Sigma_{\eta\eta} - \Sigma_{\eta\xi}\,\Sigma_{\xi\xi}^{-1}\,\Sigma_{\xi\eta}) = 0.$$

(ii) Soit $Z = (X \mid Y) \in M_{K+G,N}\,(\mathbf{R})$ un N-**échantillon aléatoire** indépendant engendré par le couple ζ précédent, avec $X = (X_1, ..., X_N)'$ et $Y = (Y_1, ..., Y_N)'$. On décompose la **matrice des covariances** empiriques S_N de Z conformément à $\zeta = (\xi, \eta)$ selon :

$$(1)_a \qquad S_N = Z\,P_{K+G}\,Z' = \begin{pmatrix} S_{XX} & S_{XY} \\ S_{YX} & S_{YY} \end{pmatrix},$$

où P_{K+G} est la matrice de centrage pr aux moyennes empiriques (cf **matrice de centrage par rapport à la moyenne**).

Les (carrés des) **coefficient de corrélation canoniques** (empiriques) sont alors les solutions $r_1^2, ..., r_K^2$ de l'équation en r^2 :

$$(2)_a \qquad \text{Dét}\,(r^2\,S_{YY} - S_{YX}\,S_{XX}^{-1}\,S_{XY}) = 0.$$

(iii) On suppose généralement que $\text{rg}\,\Sigma_{\xi\xi} = \text{rg}\,S_{XX} = K$ et que les valeurs ρ_k^2 (resp r_k^2) sont ordonnées par ordre décroissant selon $\rho_k^2 \geq \rho_{k+1}^2$ (resp $r_k^2 \geq r_{k+1}^2$), $\forall\, k \in \mathbf{N}_{K-1}^*$.